PB86-112570

Framework for Uncertainty Analysis of the NAPAP (National Acid Precipitation Assessment Program) Emissions Inventory

Brookhaven National Lab., Upton, NY

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

Sep 85

U.S. Department of Commerce
National Technical Information Service
NTIS

PB86-112570

EPA/600/7-85/036
September 1985

FRAMEWORK FOR UNCERTAINTY ANALYSIS OF THE NAPAP EMISSIONS INVENTORY

by

Carmen Benkovitz
Atmospheric Sciences Division
Department of Applied Science
Brookhaven National Laboratory
Upton, NY 11973

EPA Interagency Agreement No. DW930122-01

EPA Project Officer: J. David Mobley
Air and Energy Engineering Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

AIR AND ENERGY ENGINEERING RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA/600/7-85/036		

4. TITLE AND SUBTITLE	5. REPORT DATE
Framework for Uncertainty Analysis of the NAPAP Emissions Inventory	September 1985
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Carmen Benkovitz	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Atmospheric Sciences Division Department of Applied Science Brookhaven National Laboratory Upton, New York 11973	11. CONTRACT/GRANT NO. EPA IAG-DW930122-01
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
EPA, Office of Research and Development Air and Energy Engineering Research Laboratory Research Triangle Park, NC 27711	Final; 10/83 - 7/85 14. SPONSORING AGENCY CODE EPA/600/13

15. SUPPLEMENTARY NOTES AEERL project officer is J. David Mobley, Mail Drop 61, 919/541-2612.

16. ABSTRACT The report gives results of a project to develop a methodologies framework to assess the uncertainties associated with the emissions values as presented in the National Acid Precipitation Assessment Program (NAPAP) emissions inventory and to implement a prototype computer system to estimate the uncertainties associated with the base NAPAP emissions values. Previous projects that addressed the assignment of quantitative values to the uncertainties in emissions data were reviewed, and a statistical methodology was developed that allows the computation of more accurate uncertainty values. Auxiliary data required to carry out the computations were identified, and the prototype computer system to implement the calculations has been designed and is being implemented.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution Acidity Analyzing Computer Programs Uncertainty Principle Emission Inventories Precipitation	Pollution Control Stationary Sources Acid Rain	13B 07D 14B 09B 20J 14G 15E 04B

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release to Public	Unclassified	30
	20. SECURITY CLASS (This page) Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The objective of this project is to develop a methodologies framework to assess the uncertainties associated with the emissions values as presented in the NAPAP emissions inventory and to implement a prototype computer system to estimate the uncertainties associated with the base NAPAP emissions values. Previous projects that addressed the assignment of quantitative values to the uncertainties in emissions data were reviewed; a statistical methodology was developed that allows the computation of more accurate uncertainty values. Auxiliary data required to carry out the computations was identified and the prototype computer system to implement the calculations has been designed and is being implemented.

CONTENTS

Abstract iii
Introduction 1

1. Theoretical Background 2
   NAPAP Emissions Inventory 2
   Statistical Formulations 2
2. Application to NAPAP Emissions Inventory 5
   Yearly Emissions 5
   Temporal Resolution 7
   Spatial Resolution 8
   Chemical Speciation 9
   Spatial Aggregation 10
3. Review of Previous Work 11
4. Acquisition of Required Data 13
5. Implementation of Prototype Computer System 16
   Conceptual Design 16

References 19
Appendix - Derivation of the Error Propagation Equations 22
Acknowledgment 24

INTRODUCTION

The National Acid Precipitation Assessment Program (NAPAP) was established by Congress in 1980 to coordinate and expand research relevant to the problems associated with acid deposition in North America. The Interagency Task Force on Acid Precipitation (ITFAP) manages the program through ten subordinate task groups (Task Groups A-J) that coordinate specific technical areas of research. Task Group B coordinates research in the area of man made pollution sources.

A major goal of Task Group B is the development and maintenance of detailed inventories of anthropogenic emissions in support of acid deposition research. The Acid Deposition Emission Inventory Plan, prepared as one of the EPA contributions to Task Group B, identifies the need for emissions inventories of current base years to support both assessment activities and the development, evaluation, and use of long range transport and transformation models. The uncertainty estimates of emissions data are an integral part of the inventory. The objective of this project is to develop a methodologies framework to assess the uncertainties associated with the emissions data as presented in the NAPAP emissions inventory and to implement a prototype computer system to estimate the uncertainties associated with the base NAPAP emissions values.

SECTION 1

THEORETICAL BACKGROUND

NAPAP EMISSIONS INVENTORY

The NAPAP emissions inventory is based on the National Emissions Data System (NEDS) currently operated by the Office of Air Quality Planning and Standards (OAQPS) of EPA (National Air Data Branch, 1983). NEDS provides the basic parameters from which all other data are calculated (Engineering Science, 1984). The basic NEDS parameters allow the calculation of yearly emissions of the five criteria pollutants (particulates, SO_x, NO_x, hydrocarbons and CO) on an individual source/process basis for point sources and on a county level for area sources. To obtain the finer spatial and temporal resolution needed for some NAPAP applications, current plans call for the application of spatial, temporal and species apportionment algorithms that are based on allocation factors (or modifications thereof) developed for the Northeast Corridor Regional Modeling Project (NECRMP) (Sellers et al., 1983). Higher levels of aggregation will be calculated as sums of the NEDS data.

To establish a framework for the estimation of uncertainties associated with calculated values of the yearly emissions, this project will address the calculation of the variance of a function based on the expected values and variances of the parameters used to calculate the function. Application of these techniques will start with the algorithms used to calculate the yearly values and will be extended to include currently known algorithms for spatial aggregation and spatial, temporal and species apportionment as applicable.

STATISTICAL FORMULATIONS

The statistical equations to be used for evaluating the uncertainty of the calculated yearly emissions values are those expressing the variance of a function based on the expected values and the variances of the parameters used to calculate the function. Appendix A contains the derivation of the equations, based on probability theory, for a function x of the form

$$x = \sum_{k=1}^{N} U_k$$

and for a function x of the form

$$x = \prod_{k=1}^{N} U_k$$

To derive the values of the basic NEDS parameters widely varying methodologies have been used, ranging from statistically averaged test results to "subjective" assignment of values. These methodologies are dependent not only on the parameter in question, but also on the characteristics of the emissions source under consideration. Application of the equations derived in Appendix A assumes that a quantitative value can be assigned to the variance of the "measurement error" associated with each parameter. The derivation of these uncertainty values is to be addressed by a subsequent project.

As demonstrated in Appendix A, equations expressing the variance of a function can be greatly simplified if the parameters can be considered independent. To establish an initial framework for estimating the uncertainty of calculated emissions values, this project will assume that the "measurement errors" of the parameters in the calculating functions are independent. This assumption will be further refined as the work on estimating the uncertainties associated with the parameters is undertaken. It is anticipated that this subsequent project will examine the methodologies used to estimate the values of the parameters, identify each uncertainty associated with the use of such values - e.g. measurement error, correlations of time integrated values, etc. and evaluate how these different relationships can best be quantitatively expressed. If during the course of this future work the assumption of independent "measurement error" is proven not valid, the equations in Appendix A that include the terms with covariances and higher moments can be reexamined.

Summarizing the results for Appendix A, for N independent parameters and a function of the form $x = \sum_{k=1}^{N} U_k$ the the variance is given by

$$\boxed{V(x) = \sum_{k=1}^{N} V(U_k)} \qquad (1\text{-}1)$$

For a function of the form $x = \prod_{k=1}^{N} U_k$ the variance is given by

$$\boxed{V(x) = \prod_{i=1}^{N} \{[E(U_i)]^2 + V(U_i)\} - \prod_{i=1}^{N} [E(U_i)]^2} \qquad (1\text{-}2)$$

where E( ) represents the expected value and V( ) represents the variance of a population. Sample means (represented by u) and sample variance (represented by s^2) can be used as estimators of the population expected values and variances.

In the following sections of this report the parameter values will be assumed to represent a "sample mean value"; therefore, the bar superscript notation has been omitted in the application of equations (1) and (2).

SECTION 2

APPLICATION TO NAPAP EMISSIONS INVENTORY

YEARLY EMISSIONS

Point Sources

For each individual point source/process emissions are estimated using one of the following method:

1. Stack tests or other emission measurements
2. Material balance
3. Calculated using federal (AP-42) emission factors, where emission factors are the rate at which a pollutant is released as a result of some activity
4. Guess
5. Calculated using special emission factors.

The uncertainty of emissions estimated using method 1 can be calculated applying standard statistical equations to the original measurement data. The uncertainty of emissions estimated using method 2 must be estimated applying engineering knowledge for each process and pollutant involved. The uncertainty of emissions estimated using method 4 must be estimated based on the same criteria used to estimate the emissions value.

Emissions estimated using methods 3 and 5 are calculated using the equation:

$$EM = \left[\sum_{i=1}^{P} (EF)_i (A)_i (F)_i \right] (C) \qquad (2\text{-}1)$$

where

EM = total emissions from a point source of a given pollutant.

EF_i = emission factor for process i.

A_i = ash or sulfur content of fuel used by process i. (if appropriate)

F_i = material or fuel thruput for process i.

C = (1-control efficiency of control equipment) for a given pollutant.

P = number of processes contributing to emissions from point source.

making:

$$(EF)_i(A)_i(F)_i = (UEM)_i = \text{uncontrolled emissions of a given pollutant from process i.} \quad (2\text{-}2)$$

equation (2-1) becomes -

$$EM = \left[\sum_{i=1}^{P} (UEM)_i \right] C = (UEM)(C) \quad (2\text{-}3)$$

From equation (2-3) and using equation (1-2):

$$s_{EM}^2 = [(UEM)^2 + s_{UEM}^2] [(C)^2 + s_C^2] - (UEM)^2(C)^2 \quad (2\text{-}4)$$

From equation (2-3) and using equation (1-2):

$$s_{UEM}^2 = \sum_{i=1}^{P} s_{(UEM)_i}^2 \quad (2\text{-}5)$$

From equation (P2) and using equation (2):

$$s_{UEM_i}^2 = [(EF_i)^2 + s_{EF_i}^2] [(A_i)^2 + s_{A_i}^2] [(F_i)^2 + s_{F_i}^2] - (EF_i)^2(A_i)^2(F_i)^2 \quad (2\text{-}6)$$

Area Sources

Area source emissions for each county are calculated using the equation:

$$EM_c = \sum_{i=1}^{A} EM_a \quad (2\text{-}7)$$

where

EM_c = area source emissions for county c.

EM_a = area source emissions from activity a for county c.

A = number of activities contributing to emissions in county c.

Area source emissions for activity a are calculated using the equation:

$$EM_a = (EF)_a A_a (AC)_a \qquad (2\text{-}8)$$

where

EM_a = area source emissions from activity a in county c.

EF_a = area source emission factor for activity a (this parameter can also be county specific).

A = ash or sulfur content of fuel used in activity a (if appropriate).

AC_a = appropriate measurement of level of activity a in county c.

From equation (2-7) and using equation (1-1):

$$s^2_{EM_c} = \sum_{a=1}^{A} s^2_{EM_a} \qquad (2\text{-}9)$$

from equation (2-8) and using equation (1-2):

$$s^2_{EM_a} = [(EF_a)^2 + s^2_{EF_a}]\,[(A_a)^2 + s^2_{A_a}]\,[(AC_a)^2 + s^2_{AC_a}] - (EF_a)^2 (A_a)^2 (AC_a)^2 \qquad (2\text{-}10)$$

TEMPORAL RESOLUTION

The same temporal allocation algorithms are used for both point and area sources. Seasonal emissions for a source are calculated by:

$$EM_s = (EM_A)\,(SF)$$

where

EM_s = seasonal emissions for source.

EM_A = annual emissions for source.

SF = seasonal production fraction for source.

using equation (2)

$$s^2_{EM_s} = [(EM_A)^2 + s^2_{EM}][(SF_A)^2 + s^2_{SF}] - (EM_A)^2 (SF)^2 \quad (2\text{-}11)$$

$s^2_{EM_A}$ is obtained from equation (2-4).

Daily emissions for a source are calculated by:

$$EM_D = (EM_A)(DF)$$

where

EM_D = daily emissions for source.

DF = daily production fraction for source.

using equation (1-2):

$$s^2_{EM_D} = [(EM_A)^2 + s^2_{EM_A}][(DF)^2 + s^2_{DF}] - (EM_A)^2 (DF)^2 \quad (2\text{-}12)$$

Hourly emissions for a source are calculated by:

$$EM_H = (EM_A)(HF)$$

where

EM_H = hourly emissions for source.

HF = hourly production fraction for source.

Using equation (1-2)

$$s^2_{EM_H} = [(EM_A)^2 + s^2_{EM_A}][(HF)^2 + s^2_{HF}] - (EM_A)^2 (HF)^2 \quad (2\text{-}13)$$

SPATIAL RESOLUTION

Spatial allocation factors need only be developed for area sources, since point source information includes specific location of the source.

Spatial allocation of area sources to grid cells is accomplished by

$$EM_i = \sum_{c=1}^{C} \sum_{a=1}^{A} (EM_{ac})\,(AF_{ac})$$

where

EM_i = total emissions in grid cell i.

EM_{ac} = emissions factor for activity a in county c.

AF_{ac} = allocation factor for activity a in county c.

C = number of counties whose area is completely or partially in grid cell i.

A = number of activities in county c

Making

$B_{ac} = (EM_{ac})\,(AF_{ac})$ and using equation (1-2):

$$s^2_{B_{ac}} = [(EM_{ac})^2 + s^2_{EM_{ac}}][(AF_{ac})^2 + s^2_{AF_{ac}}] - (EM_{ac})^2\,(AF_{ac})^2 \qquad (2\text{-}14)$$

$s^2_{EM_{ac}}$ is obtained from equation (2-10)

$$s^2_{EM_i} = \sum_{c=1}^{C} \sum_{a=1}^{A} s^2_{B_{ca}} \qquad (2\text{-}15)$$

CHEMICAL SPECIATION

Speciation of SO_x emissions to SO_2/SO_4, of NO_x emissions to NO/NO_2 and of hydrocarbons to specific chemical species will be accomplished via the use of speciation factors expressed as weight percent of uncontrolled emissions.

Therefore, equation (2-3) becomes:

$$EM = \left[\sum_{i=1}^{P} (UEM)_i\,(SP)_i\right] C = (UEM)\,(C)$$

EM = total emissions from a point source of a given pollutant.

$(UEM)_i$ = uncontrolled emissions from process i of a given pollutant.

$(SP)_i$ = speciation factor for process i of a given species.

C = (1-control efficiency of control equipment) for a given pollutant.

Equation (2-5) then becomes:

$$s^2_{UEM} = \sum_{i=1}^{P} \{[(UEM_i)^2 + s^2_{UEM_i}] [(SP_i)^2 + s^2_{SP_i}] - (UEM_i)^2 (SP_i)^2\} \quad (2\text{-}16)$$

where $s^2_{UEM_i}$ is obtained from equation (2-6).

$$s^2_{EM} = [(UEM)^2 + s^2_{UEM}][(C)^2 + s^2_c] - (UEM)^2 (C)^2 \quad (2\text{-}17)$$

SPATIAL AGGREGATION

All spatial aggregation calculations start with emissions at the point source or county level. All are computed as sums of these basic levels. Therefore the variance of these aggregated emissions are calculated as the sum of the variances of the individual components.

SECTION 3

REVIEW OF PREVIOUS WORK

Several previous projects have addressed the problem of assigning, in statistical terms, quantitative values to the errors in emissions data. Final reports from the following projects were reviewed.

1. Weighted Sensitivity Analysis of Emissions Data, project conducted by IBM Corporation under contract to the Office of Air Quality Planning and Standards (OAQPS), EPA, July 1973.
2. Source Inventory and Emission Factor Analysis (SIEFA) project conducted by PEDCo - Environmental Specialists, Inc. under contract to OAQPS, September 1974.
3. Emissions Inventory for the SURE Region, project conducted by GCA Corporation under contract to the Electric Power Research Institute (EPRI), April 1981.
4. Emissions, Costs and Engineering Assessment, Work Group 3B, US-Canada Memorandum of Intent (MOI) on Transboundary Air Pollution, June 1982.
5. Preliminary Evaluation of Acidic Deposition Assessment Uncertainties, project conducted by Argonne National Laboratory (ANL) under contract to the U.S. Department of Energy, November 1982 (preliminary report).

All these projects have based their calculations on the statistical formulas for error propagation as derived for "small" values of the errors; i.e., the Taylor series expansion included only the first derivatives of the function. Some of the error values used in subsequent calculations have been as high as 80-90% of the mean. For this size errors the exact equations as derived in the Appendix of this report, represent more accurate solutions.

The Weighted Sensitivity Analysis project used a top down approach to the calculation of errors in emissions data; i.e., for a maximum desired error at the top of a hierarchical structure, what are the maximum errors that can be allowed in the basic quantities? All other projects employed a bottom up approach, assigning error values to a set of basic quantities and calculating the resulting error at the top of a hierarchical structure.

The SIEFA project was the first to address the assigning of error values to the NEDS basic levels, i.e., at the individual source level for point sources and at the county level for area sources. The two volume

final report details the data and calculations used to derive the precision values used in this study. Precision was defined at

$$P_i = \frac{S_i}{\bar{X}_i}$$

where

P_i = precision of component i

S_i = standard deviation of the mean of component i

$\bar{X}_i$ = arithmetic mean of component i

Derivation of the precision values of the individual components of the area source calculations were developed based on the methodology used by NEDS in 1974; most of these methodologies are not currently valid.

The SURE emissions inventory project developed emissions figures for an area that included the eastern United States and southeastern Canada. Chapter 5 of the final report describes the methodology used to evaluate the accuracy of the resulting inventory. The estimation of errors associated with the emission estimates was addressed for five source classifications. The approach used to estimate the errors in any one category was similar to the methods used in the SIEFA project; error factors from SIEFA were used as one of the data sources. In the report component errors are tabulated in percent by pollutant and source categories; no details are given on how the error figures were derived.

The MOI project report includes a sample table of the "probable errors" (in percent) derived for each source category and for each state. Source categories used in this project do not match those developed for the SURE project; no details are given on how the error figures were derived.

The ANL project was entirely based on the error analysis techniques and data derived for the SURE project.

SECTION 4

ACQUISITION OF REQUIRED DATA

Information currently available in the NAPAP emissions inventory includes data on yearly emissions of point sources at the individual source level and on yearly emissions of area sources at the county/category level. Point source emissions values have been estimated either from stack tests, material balance or engineering judgment or have been calculated using federal (AP-42, see references) or local emission factors. For the calculated values, emission factors are applied at the individual process level. Source Classification Codes (SCC) are used to identify the processes and point source emissions are calculated as the sum of the emissions of all the individual processes contributing to the source. Area source emissions are calculated using the appropriate emission factors. Parameters used to calculate emissions values include emission factors, fuel, process or activity rates, ash or sulfur content of fuels (if appropriate) and efficiency of control equipment for the appropriate point sources or category adjustment factors for area sources.

As previously discussed in the Theoretical Background section of this report, the basic NEDS data have been derived using a wide variety of different methodologies and data sources. These differences are not only dependent on the parameters under consideration, but may also depend on the characteristics of each emissions source.

The development of quantitative values for the uncertainty of emissions estimates or of the parameters needed for the calculation of the emissions values is to be addressed by a subsequent project. Additional data to be developed are as follows:

- Uncertainty of emissions estimated using stack tests, material balances or engineering judgment. The data and methodologies used to estimate the emissions values must be studied to select the appropriate statistical techniques that must be applied to calculate the uncertainty of these emissions values.

- Uncertainty of federal and local emission factors. The data and methodologies used to calculate emission factor values must be studied to select the appropriate statistical techniques that must be applied to calculate the uncertainty of these values. Additional topics to be examined must include the uncertainties introduced by

the application of the derived emission factors to the inventory population.

- Uncertainty of ash and sulfur contents of fuels: coals, oil, gas.

- Uncertainty of point source fuel or process rates. A literature survey to determine how the currently used values were derived and applied would provide a starting point for estimating the values of these uncertainties Panels of experts in these fields should be consulted to refine these estimates.

- Uncertainty of efficiency of control equipment. This is an area where a lot of controversy exists. Existing EPA data on control equipment testing could provide a starting point for estimating these variances. Experts in the field should be consulted to refine preliminary estimates; care must be exercised to insure that dependencies between control efficiencies and other parameters are properly identified and resolved within the framework of the uncertainty project methodology.

- Uncertainty of area source activity rates.

- Uncertainty of area source category adjustment factors. These rates and adjustment factors are generally estimated using data gathered by other agencies, i.e., census data, DOT data on motor vehicle use, etc. EPA personnel working in the field of area source emissions calculations should be consulted to determine what data are used and where these data originate. The appropriate agencies may have to be contacted to provide these uncertainty values or to determine the best way to estimate those values needed.

Table 1 (composite data from the Engineering-Science report, see references) provides a starting point for the selection of priorities in the acquisition of the uncertainty values needed. However, further breakdowns of the values presented in these figures and tables may be needed to fully define the priority list.

TABLE 1. NAPAP EMISSIONS INVENTORY - 1980 BASE YEAR*

	SO_x	NO_x	HC(1)
Total emissions (10^6 tons)	27.1	23.7	23.3
Point sources (percent)	92.2	48.9	21.3
Area sources (percent)	7.8	51.1	78.7
Electric utilities (percent)	63.9	34.2	--
Highway vehicles - gasoline (percent)	--	21.2	29.0
Miscellaneous (percent)	--	--	41.2
Diesel fuel use (percent)	--	16.1	2.7

* Source: Toothman et al., EPA-600/7-84-091.

Temporal resolution of the yearly emissions is obtained by the use of equations T1-T3. Spatial apportionment of area source emissions is accomplished via the use of allocation factors for each activity in each county. Chemical speciation of the yearly emissions is obtained by the use of speciation factors keyed to SCC codes. The apportionment calculations are scheduled to be implemented as add-on capabilities to the basic NAPAP emission inventory system. Allocation factors that were developed for other emissions inventory projects are being studied and adapted for use in the NAPAP inventory. Uncertainty values for all allocation factors and the calculation of the uncertainty associated with the allocation calculations are within the scope of these projects.

SECTION 5

IMPLEMENTATION OF PROTOTYPE COMPUTER SYSTEM

This section addresses the design and implementation of the prototype computer software needed to calculate uncertainties associated with yearly emissions values for both point and area sources. The conceptual design is independent of the software system used to support the NAPAP inventory; the design and implementation of the software modules will allow portability between computer systems is as independent as possible of the current NAPAP inventory software system.

Calculation of the uncertainties associated with the temporal, spatial and species apportionment of emissions values are to be based on the methodology developed in this project but will be included in the software system designed to calculate these values.

CONCEPTUAL DESIGN

Uncertainty values associated with each parameter used in calculating emission values are expected to be applicable for whole "classes" of these parameters rather than just to individual sources. To avoid needless repetition of data (with the associated increase in storage costs and the ever present danger at update times), a file with uncertainty "profiles" will be created. This file will contain all the information needed to calculate the uncertainty associated with the yearly emissions of a point source or an area source category. Each uncertainty profile will be identified by a unique code; at retrieval time the appropriate uncertainty profiles can be referenced to allow the calculation of uncertainty values associated with the reported emissions. Figure 1 presents a schematic of this design.

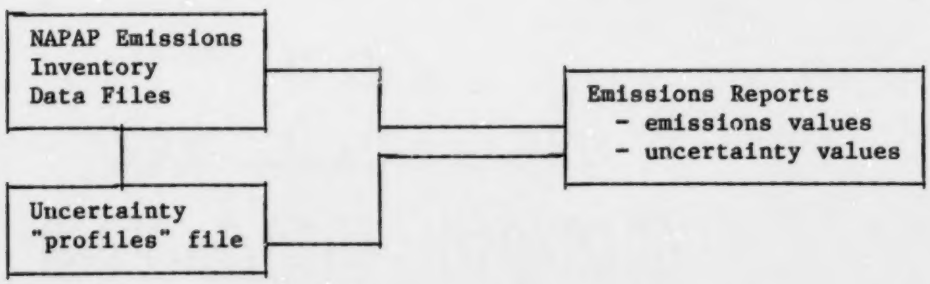

Figure 1. Implementation Based on the Current NAPAP Emissions Inventory System

Contents of the uncertainty profiles will only be defined in a generalized format. Defining the specific contents of such profiles is part of the scope of work of the project developing the uncertainty values of the basic calculation parameters.

The current version of the NAPAP emissions inventory has been implemented based on the Emission Inventory System (EIS) on the EPA IBM computer at Research Triangle Park, N.C. EIS is comprised of a body of programs written in COBOL that create and maintain the data files needed to store inventory information, and access and manipulate these data files to retrieve needed information to produce the desired reports. References by McMaster present detailed descriptions of EIS data files and retrieval programs.

Due to the rigid structure of EIS, modification of this system to accommodate the uncertainty calculations methodology and data are being kept to a minimum. Only currently unused space in EIS data files is being used for needed pointers; all the data needed for uncertainty calculations is stored outside EIS. Programs needed to maintain the uncertainty pointers and flags in the EIS files and to calculate uncertainty values are being implemented as modular programs using the standard FORTRAN 77 language.

Implementation of the conceptual design for the current version of the NAPAP inventory includes the following tasks:

1. Point sources whose emissions values are estimated from stack tests, material balance or engineering judgement have no valid data in the EIS/PS Source Classification Code (SCC) record; therefore, any uncertainty information available for these sources must be linked to the EIS/PS point record level. Point sources whose emissions values have been estimated using emissions factors do have valid data in the EIS/PS SCC record; uncertainty data for all the pertinent parameters used in the calculations (emissions factors, ash/S content of fuel, process rates, etc.) will be linked to this EIS/PS record. For area sources uncertainty data for all the pertinent parameters (emissions factors, ash/S content of fuel, level of activity, etc.) will be linked to the EIS/AS county/category level. At the point record level the uncertainty profile ids are stored in characters 177-185 of the EIS/PS Master file-point record; for the SCC level these ids are stored in characters 135-143 of the EIS/PS Master File-SCC Record (these fields are currently unused for the NAPAP data). For area sources the uncertainty profile ids are stored in characters 252-260 of the EIS/AS Master File-Category Record. A nine character id allows for 362,880 uncertainty profiles; in terms of possible SCC values this allows nine profiles per SCC. Software is being developed create and maintain the uncertain profile id fields in the EIS/PS and EIS/AS master files.

2. A generalized format for the uncertainty profiles file is being defined. Each profile will include all the data needed to calculate the uncertainty of the values of the yearly emissions of the point source or area source category with whom this profile is associated. Detailed implementation will be carried out by a subsequent project.

3. Modular software is being written to apply the uncertainty equations to individual point sources (equations P4-P6), to county level area sources (equations A3-A4) and spatial aggregations of both.

4. Reports are being developed to allow users to summarize emissions estimates and calculate associated uncertainties in one of five formats for point sources: point, plant, AQCR, county or state and in one of three formats for area sources: AQCR, county or state.

REFERENCES

Beers, Y. Introduction to the Theory of Error. Addison-Wesley Publishing Co., Inc., 1957.

Bevington, P.R. Data Reduction and Error Analysis for the Physical Sciences, McGraw Hill, 1969.

Ditto, F.H., L.T. Gutierrez, T.H. Lewis and L.J. Rusbrook. Weighted Sensitivity Analysis of Emissions Data: Volume I and II. EPA-450/3-74-022 a and b, 1973.

Eisenhart, C. Expression of the Uncertainties of Final Results. Science 160:1201-1204, 1968.

Klimm, H.A. and R.J. Brennan. Emissions Inventory in the SURE Region. EPRI-EA-1913, 1981.

Knudson, D., M. Davis, J. Shannon, D. Sisterson, S. Viessman, M. Wisely and R. Whitfield. Preliminary Evaluation of Acidic Deposition Assessment Uncertainties. ANL/EES-TM-XXX (Draft), 1982.

McMaster, L.R. The Emissions Inventory System/Area Source User's Guide. EPA-450/4-80-009, 1980.

McMaster, L.R. The Emissions Inventory System/Point Source User's Guide. EPA-450/4-80-010, 1980.

Meyer, P.L. Introductory Probability and Statistical Applications. Addison-Wesley Publishing Co., 1970.

National Air Data Branch, Environmental Protection Agency, Final Work Plan-Acid Deposition Emission Inventory Data Base Management and Maintenance. Contract No. 68-02-3509, Engineering Science, Fairfax, Va., 1983.

Office of Air Quality Planning and Standards, Environmental Protection Agency. Source Inventory and Emission Factor Analysis. Volumes I and II, EPA-450/3-75-082 a and b, 1974.

PEDCo - Environmental Specialists, Inc. Source Inventory and Emission Factor Analysis, Volumes I and II. EPA-450/3-75-082 a and b, 1974.

Rivers, M.E. and K.W. Riegel (Eds.). Emissions, Costs and Engineering Assessment Work Group 3B. United States-Canada Memorandum of Intent on Transboundary Air Pollution, 1982.

Robbins, H. and J.U. Ryzin. Introduction to Statistics. Science Research Associates, Inc., 1975.

Sellers, T.M., M.J. Giraghty, A.M. Kiddu, B.J. Bosy and S.U. Capone. Northeast Corridor Regional Modeling Project Annual Emissions Inventory Compilation and Formatting, Volume XVII, Development of Temporal, Spatial and Species Allocation Factors. EPA-450/4-82-013q, 1983.

Sellers, T.M. Allocation Factors and Algorithms for the National Acid Precipitation Assessment Program Emissions Inventory. Technical Memorandum GCA-TR-83-65-G, GCA Corporation, GCA/Technology Division, 1983.

Snedecor, G.W. and W.G. Cochran. Statistical Methods. Iowa State University Press, 1967.

Toothman, D.A., Yates, J.C. and Sabo, E.J. Status Report on the Development of the NAPAP Emission Inventory for the 1980 Base Year and Summary of Preliminary Data. EPA-600/7-84-091, 1984.

APPENDIX

Derivation of the Error Propagation Equations

A) For a function of the form:

$$X = \sum_{k-1}^{N} U_k$$

then

$$V(X) = E(X^2) - [E(X)]^2$$

$$= E\left[\sum_{i,j} U_i U_j\right] - \left[\sum_{i} E(U_i)\right]^2$$

$$= \sum_{i,j} E(U_i U_j) - \sum_{i,j} E(U_i)E(U_j)$$

$$= \sum_{i,j} \text{cov}\,(U_i U_j)$$

$$V(X) = \sum_{k=1}^{N} V(U_k) + 2 \sum_{k=1}^{N} \sum_{j=k+1}^{N} \text{cov}\,(U_j U_k)$$

If the U_k are independent variables this equation reduces to

$$V(X) = \sum_{k-1}^{N} V(U_k)$$

B) For a function of the form

$$X = UW$$

make $\mu = E(U)$ $\quad \nu = E(W)$

$U_0 = U - \mu$ $\quad W_0 = W - \nu$

$\rho = \text{cov}\,(U,W)$

then

$E(U_0) = 0$ $\quad E(W) = 0$

$E(U_0^2) = V(U)$ $\quad E(W^2) = V(W)$

$$V(X) = E(U_0 + \mu)^2 (W_0 + \nu)^2 - [E(U_0 + \mu)(W_0 + \nu)]^2$$

$$= E(U_0^2 W_0^2) + \mu^2 V(W) + \nu^2 V(U) + \nu^2 \mu^2$$

$$+ 2\mu E U_0(W_0^2 + \nu^2) + 2\nu E W_0(U_0^2 + \mu^2) + 4\mu\nu\rho - (\rho + \mu\nu)^2$$

$$= \nu^2 V(U) + \mu^2 V(W) + 2\mu\nu\rho + 2\nu E(U_0^2 W_0) + 2\mu E(W_0^2 U_0) + E(U_0^2 W_0^2) - \rho^2$$

$$= \nu^2 V(U) + \mu^2 V(W) + 2\mu\nu \, Cov(U,W) + 2\nu E(U-\mu)^2 (W-\nu)$$

$$+ 2\mu E(U-\mu)(W-\nu)^2 + E(U-\mu)^2 (W-\nu)^2 - [Cov(U,W)]^2$$

If U and W are independent variables this equation reduces to

$$V(X) = \nu^2 V(U) + \mu^2 V(W) + V(U) V(W).$$

Extending the above equation to a function of the form

$$X = \prod_{k=1}^{N} U_k$$

where the U_k are independent variables

$$V(X) = \prod_{i=1}^{N} EU_i^2 - \left(\prod_{i=1}^{N} E U_i \right)^2$$

$$= \prod_{i=1}^{N} \{[E(U_i)]^2 + V(U_i)\} - \prod_{i=1}^{N} [E(U_i)]^2$$

ACKNOWLEDGMENT

The author wishes to acknowledge the contributions of Neal Oden to the development of the statistical basis of this report and of Paul Michael for serving as a general consultant and interested party in both the scientific and administrative aspects of this project.